COPY NO.

IMM-NYU 207
MARCH 1954

NEW YORK UNIVERSITY
INSTITUTE OF MATHEMATICAL SCIENCES
LIBRARY
25 Waverly Place, New York 3, N. Y.

IMM207

NEW YORK UNIVERSITY
INSTITUTE OF
MATHEMATICAL SCIENCES

Lectures on the Theory of Industrial Sampling

J. H. CURTISS

PREFACE AND SUMMARY OF THE TEXT

TECHNICAL REPORT II

CONTRACTS NOS. DA-30-069-ORD-1257
AND
N6ori-201, TASK ORDER NO. 1
PREPARED UNDER THE SPONSORSHIP OF
ARMY OFFICE OF ORDNANCE RESEARCH
AND
OFFICE OF NAVAL RESEARCH

c.1

IMM-NYU 207
March 1954

LECTURES ON THE THEORY OF INDUSTRIAL SAMPLING

BY

J. H. CURTISS

Notes for a graduate course given at New York University in the fall term of the academic year 1953-1954.

Preface and summary of the text.

Note: The purpose of this technical report is to provide information about a longer work now in rough draft form, with the idea of determining processing requirements and obtaining suggestions for improvement.

Contracts Nos. DA-30-069-ORD-1257

and

N6ori-201, Task Order No. 1

Technical Report No. 2

This report represents results obtained at the Institute of Mathematical Sciences, New York University, under the sponsorship of the Army Office of Ordnance Research and the Office of Naval Research.

New York, 1954

Lectures on The Theory of Industrial Sampling

by J. H. Curtiss

Preface. The purpose of these lectures is to lay the theoretical groundwork for a full understanding of industrial sampling procedures. Only probability sampling is covered, and not purposive or judgment sampling

The emphasis throughout is primarily on theory and on rationale, and not on practice. It is true that in the later chapters, a number of specific procedures for setting up acceptance sampling plans are spelled out in some detail. But this is not done with the idea of making these Notes useful as an engineers' manual of quality control methods. Instead, the purpose is to provide the reader with concrete illustrations which will aid him in his own applications of the theory.

For the most part, no attempt was made to reproduce in the lectures proofs of standard results in distribution theory. In the first place, most of the important fundamental material of this type is now readily available in the text-book literature. (Cramer's Mathematical Methods of Statistics, Feller's Probability Theory, and Kendall's The Advanced Theory of Statistics cover much of the field.) In the second place, it is the author's opinion that very few, if any, of the more complicated standard proofs in mathematical statistics really give much insight into the meaning of the results. An exception is the Neyman-Pearson Theorem in Chapter V, for which a proof is accordingly supplied.

The only chapter which is fundamentally affected by this attitude toward the classical proofs is the basic chapter (Chapter IV) on distribution theory. The presentation there takes the form of a compendium of results rather than that of a classroom presentation. The advantage of this approach is of course that a good deal of territory can be compressed into a relatively small amount of space. For example, Chapter IV includes a rather complete treatment of the distribution of linear and quadratic forms in independent normally distributed random variables. It is hoped that the collection of results in this chapter will be found useful as reference material for various studies in mathematical statistics in addition to the present one.

CHAPTER I

FOUNDATIONS OF MATHEMATICAL PROBABILITY

Section 1. The undefined elements of the theory.

Section 2. Composite events.

Section 3. Algebra of events.

Section 4. Probability.

Section 5. Conditional probability.

Section 6. Independence.

Section 7. Random variables (r.v.)

CHAPTER II

MULTIVARIATE CONCEPTS AND TERMINOLOGY

Section 1. Joint distributions. This topic is approached by defining the joint distribution of several r.v.'s Y, Z, U, ... which are all functions of the same r.v. X.

Section 2. Marginal distributions.

Section 3. Conditional distributions and independence.

Section 4. Special case: the components of a vector r.v.

Section 5. Repeated trials and composite experiments. Bernoulli trials are introduced here, and the binomial distribution. Also the "chain method" of specifying the probability distribution (p.d.) of a sequence of dependent r.v.'s.

CHAPTER III

FUNDAMENTALS OF SAMPLING THEORY

Section 1. The basic sampling experiment. Sampling from a finite population or lot is introduced, together with the appropriate terminology. The event space for the selection operation is carefully distinguished from that for the measuring and classifying operation involved in a sampling experiment.

Section 2. The empirical distribution and purposive sampling. The empirical frequency distribution of the sample is introduced.

Section 3. Probability sampling. Definitions are given of probability sampling in general, and of random sampling in particular. It is shown that in random sampling, the marginal p.d. of each observation (treated as a r.v.) reproduces the true frequency distribution of the lot. Random sampling with and without replacements are distinguished. It is noted that any uniform joint distribution on a rectangular n-dimensional region necessarily implies statistical independence of components.

Section 4. Extensions of the sampling concept and terminology. The concept of random sampling from an "infinite" lot is introduced by stages.

Section 5. Sampling from a finite population to which an approximate frequency distribution has been fitted. Extension of Section 4.

Section 6. Distribution functions (d.f.). The main properties

of d.f.'s are introduced without proofs. There is a discussion of the error in a joint p.d. when the p.d.'s of the components are subject to error (extension of Section 5).

Section 7. Relations between the empirical distribution and the frequency distribution of the lot. The empirical d.f. is introduced. The facts concerning convergence (Glivenko-Cantelli Theorem) of the empirical d.f. to the lot frequency d.f. under random sampling are stated carefully, but without proofs. Finally the generalization to place selections which forms the characterization of "randomness" contained within the present approach to probability theory, is stated but not proved. Practical implications of these theorems are noted. An example of a uniformly distributed non-random sequence is given.

Section 8. Random selection versus non-random selection. A discussion of the advantages of probability sampling in the light of the theorems of Section 7. The discussion centers on a simple example.

Section 9. On insuring that sampling is random. Theory and use of tables of random sampling numbers. It is proved that if the digits of a multi-digit number have a uniform distribution and are independently distributed, then this property carries over to the number itself. Various conditional distributions, relative to rejection of blocks of numbers in a table of random numbers, are established, and their application is described. A sequence of deterministic pseudo-random numbers useful for "Monte Carlo" work (model-sampling) is given.

CHAPTER IV

SOME DISTRIBUTION THEORY

Section 1. *Frequency vs. probability distributions.* It is pointed out that within the present approach to probability, all of the *mathematical* apparatus developed for probability distributions applies to relative frequency distributions as well.

Section 2. *Specification of distributions.* The distinction between discrete and continuous distribution, and the concept of density function, are introduced.

Section 3. *Expected values.* These are defined in general, and the main theorems are stated without proof. The fact that an expected value of a r.v. is the almost certain limit of the arithmetic mean of a sequence of independent determinations of the r.v. is stated to provide a motivation for the alternative terminology, "mean value".

Section 4. *Moments.* The standard definitions and inequalities are given. The moments of the empirical distributions are introduced, and the mean and variance of X and s^2 in random sampling are calculated to illustrate the theorem of Section 3.

Section 5. *Moment generating functions and characteristic functions (c.f.).* Definitions, relation to moments. The inversion and continuity theorems are stated accurately, but without proofs. The c.f. of a sum of independent r.v.'s is noted.

Section 6. *Central Limit Theorems.* Definition of the normal or Gaussian distribution. The theorem of Liapounov is stated, with indication of proof. The Berry-Eseen estimate of

error is given in a simplified form.

Section 7. *Convergence theorems for distribution of a sum of random binary digits*. Various forms of the DeMoivre-Laplace Theorem are stated.

Section 8. *Convergence to a uniform distribution*. The Levy theorem regarding convergence of a sum of independent r.v.'s reduced modulo the unit interval, to the uniform distribution, is stated and the proof is indicated. Two results concerned with "purifying" random binary digits are given as exercises.

Section 9. *Some properties of special distributions*. A methodical summary of the properties of several important distributions is given. The facts for each distribution are arranged under the headings of *Formulas for the d.f.*, *C.f. and moments*, *Stability* (under addition of r.v.'s), *Properties of special functions*, *Tables*. The distributions treated are the binomial, hypergeometric, Poisson, and normal. The arcsine and square-root transformations for the binomial and Poisson are included. In the case of the normal distribution, the basic facts about linear and quadratic forms in independent one-dimensional normal variables are stated. The Chi-Square distribution is introduced. The Fisher-Cochran Theorem is given, together with algebraic results to facilitate its use. The A. T. Craig Theorem giving a necessary and sufficient condition for two quadratic forms to be independent in terms of the product of their matrices, is stated. The "t" and "F" distributions are introduced, together with

exercises relating the latter to the distribution of "z" and to the binomial distribution. The Chi-square distribution is related to the Poisson.

CHAPTER V

INTRODUCTION TO THE THEORY OF ACCEPTANCE SAMPLING

Section 1. Definition of acceptance sampling. Measures of item quality are distinguished from measures of lot quality, and acceptance sampling is defined in terms of these measures.

Section 2. Techniques of statistical inference useful in acceptance sampling work. Point estimation, interval estimation, and the testing of statistical hypotheses are defined mathematically and contrasted, and reasons are given for preferring the third technique in acceptance sampling work. (The definition of tests of statistical hypotheses covers randomized tests of parametric hypotheses consisting of multiple alternatives; in other words, a fairly general case of the decision problem is described.)

Section 3. Acceptance sampling based on tests of statistical hypotheses. This section contains a careful formulation of the classical Neyman-Pearson theory of tests of parametric hypotheses in the language of industrial sampling theory. It is pointed out that the formulation leaves open various questions relating to the specification of lot distributions, optimum choice of acceptance regions, and selection of the regions in the parameter space representing alternative hypotheses.

Section 4. The operating characteristic (o.c.), and consumers' and producers' risks. These are defined in a very general way.* A preliminary discussion of what constitutes a

---

*Non-parametric definitions of consumers' and producers' risks are given which may be new.

good test is given in terms of the definitions. Graphs of typical o.c.'s are given.

Section 5. *Illustrations*. *Lot quality measured by the proportion not conforming to a standard*; *lot quality measured by the mean of a normal distribution with bounded standard deviation*; *lot quality measured by the standard deviation of a normal distribution*; *lot quality measured by the median of an arbitrary distribution*. First a detailed general outline of the information and calculations necessary to set up a decision-type sampling plan is written out. Then special sampling plans for each of the hypotheses listed in the title of the section are presented. The discussion in each case follows the general outline strictly. In the binomial case, the use of the Tukey-Mosteller approximation to the significance levels of the binomial distribution is demonstrated and the corresponding formulas for acceptance numbers and sample size are given. In the continuous cases, the importance of taking into account the *scale of measurement* in defining good and bad quality is emphasized. There is an elaborate discussion with a figure in the normal case (quality measured by the mean) as to why it is legitimate to use a known *bound* on the lot standard deviation in the formulas* instead of a known fixed value of the lot standard deviation. The question of the power of the median test is carefully treated, and a restriction on the hypothesis is proposed which allows an o.c. to exist.

---

* New.

Section 6. A priori information regarding the frequency distribution of the lot.* This section deals with the question of the validity of specifying simple standard types of frequency distributions as lot distributions, on which to base acceptance sampling plans. Three approaches are distinguished: the engineering approach, the probabilistic approach, and the empirical approach. In the course of the discussion of the second one, special models are presented leading to the normal and logarithmo-normal distribution and Poisson distribution. Specification of lot distributions based on a large sample theory of the lot is distinguished from lot distribution specification based on a conditional distribution theory of the sample. (In the latter, the lot is viewed as a sample from a larger population, the process.)

Section 7. The optimum choice of the acceptance set. An example is given in which intuition would probably lead to a very bad acceptance region. Then the Neyman-Pearson Theorem regarding likelihood ratio tests for simple alternatives is stated and its implications are discussed. (Proof given in Appendix to the Chapter.)

Section 8. Illustration of likelihood ratio tests. The likelihood ratio test region is calculated for tests of simple hypotheses regarding parameters of the binomial distribution, the Cauchy distribution, and the normal distribution (a general discussion is given in this case covering all possible simple hypotheses on the mean and standard deviation).

---

*A connected discussion of the subject matter of this section does not seem to have been previously available.

Section 9. Optimum acceptance regions in the case of composite hypotheses. The difficulties of arriving at optimum tests for composite hypotheses are briefly discussed. Then a procedure is given for extending an UMP test for a simple null hypothesis vs. a composite alternative hypothesis, to take care of a composite null hypothesis.*

Section 10. Effect of errors of measurement and classification on the o.c.. The modification of an o.c. introduced by errors of measurement is first discussed in full generality. The specialization to a stationary distribution of errors is then treated, and is reformulated in terms of a sum of r.v.'s. The certain decrease in the sharpness of the o.c. in this specialization is pointed out. Then the dichotomous case is considered in detail (sampling with replacements) and a formula is derived which identifies the exact quality level at which the decrease in sharpness of the o.c. begins in either direction.

Appendix to Chapter V. (This contains detailed mathematical arguments postponed from the text.) Results of Mosteller-Tukey and Freeman-Tukey regarding approximations to the significance levels of the binomial distribution are discussed. The distribution of the k-th order-statistic of a sample is derived. The Neyman-Pearson Theorem is proved.

---

*New

CHAPTER VI

SAMPLING FROM A LOT WHICH ITSELF IS A PROBABILITY SAMPLE FROM A PROCESS *

Section 1. *Introduction*. It is pointed out that the chapter is concerned with a situation already described in Chapter V in connection with the *probability approach* to the problem of specifying a frequency distribution for the sampled lot.

Section 2. *Lot acceptance based on the conditional distribution theory of the sample*. Two goals of lot acceptance sampling when the lot is viewed as a sample are formulated. (They are *process acceptance*, and *lot acceptance*.) The technique of constructing a conditional lot quality decision procedure based on the hypothesis that a certain r.v. Θ^* representing lot quality has a fixed value is developed. The meaning of the resulting conditional o.c. is carefully contrasted with the meaning of the o.c. in the fixed lot case.

Section 3. *Classification of consumers' and producers' risks*. Statistical control is defined. Consumers' and producers' lot-risks and process-risks are defined.

Section 4. *Sampling from a lot which itself is a random sample*. A basic theorem is stated and proved which states in effect that if a lot is a *random* sample *with* replacements from a process, then any sample (random or not) extracted from the lot in any way *without replacements* is a random sample from the process. Under these hypotheses, the uninspected remainder

---

*The material in this Chapter is largely new, and will be submitted separately for publication to the Journal of the American Statistical Association. Abstract submitted to Institute of Mathematical Statistics for presentation by title at Summer 1954 meeting.

of the lot is always distributed independently of the sample from the lot.

Section 5. Rationale of lot acceptance sampling when the sample and the uninspected remainder of the lot are statistically independent. The discussion first cites possible legal and contractual motives for sentencing on overall lot quality rather than on the quality of the uninspected remainder. It then deals at length with the non-statistical inferences which can be drawn from one of two independent samples concerning the other (here the independent samples are the samples from the lot and the uninspected remainder). Various numerical inequalities which connect parameters of the lot, of the sample, and of the uninspected remainder of the lot, are derived.

Section 6. Some conditional distributions and sampling plans based on them. This section develops some of the distribution theory needed to set up sampling plans based on conditional o.c.'s relative to hypotheses concerning parameters of a lot which itself is a random sample. A general theorem is first proved which states in particular that the conditional mean value of any sample moment about a fixed point, given that the corresponding lot moment is equal to K, is also equal to K. Then the conditional distributions of the sample percent defective, and of the sample mean, and of the sample standard deviation, corresponding to hypotheses on the corresponding parameters of the lot, are derived. (The assumption of a normal process is made in the last two instances.) The conditional distribution of the total

number of successes in a sample from a larger sample of Poisson r.v.'s is also derived. The conditional o.c. for conventional hypotheses as to what constitutes good and bad quality are set up in each case.

Section 7. *Simultaneous use of L-risks and P-risks*. It is shown how the constants of certain sampling plans can be determined by imposing one condition on the unconditional o.c. and another condition on the conditional o.c. relative to a hypothesis on lot quality. The possibility of using one quality measure for the process and another quite different one for the lot is noted.

Section 8. *Some questions concerning the average behavior of lot acceptance sampling plans based on the conditional distribution theory of the sample*. The following questions are answered by formally performing the appropriate summations on a general conditional o.c.: (1) What is the unconditional probability of accepting a bad lot? Is it as large as the consumers' L-risk α_L ? (2) What is the conditional probability of accepting a lot, given that it is bad? Is it less than α_L? (3) What is the conditional probability distribution of the quality of the uninspected remainder of the lot, given that the lot is accepted? (4) Does lot acceptance under present circumstances improve the quality of the accepted product, assuming that the process remains in a fixed state of control? (The answer is yes.)

CHAPTER VII

FURTHER SAMPLING PLANS

This chapter has not as yet been written up. The presently available rough lecture notes include the theory and application of the sample range in acceptance sampling; theory and use of the mean range; comparison of efficiency of range, mean range, and sample standard deviation; range - mid-range tests on the median of a normal distribution; control of percent nonconforming by variables inspection. Another topic which would be included in a final write-up would be sampling plans intended to test whether or not the lot frequency distribution adequately conforms to a preassigned standard frequency distribution (based on the Chi-square and Kolmogorov-Smirnov tests).

Before formal publication of these notes, the following chapters would be added to the above material:

CHAPTER VIII

PROCESS CONTROL AND CONTINUOUS INSPECTION PLANS

This chapter would be concerned with sampling plans which are applied repeatedly or continuously to improve or check on process quality. There would be two distinct parts to the chapter. In the first part, the theory of the control chart would be developed. There would be a discussion of the power of the significance tests, involved in control chart techniques. Nonparametric tests of randomness and trend useful in connection with control charts would be included. In the second part, AOQ and

AOQL would be defined and applied to acceptance sampling. Then the Wald-Wolfowitz continuous inspection scheme would be presented.

CHAPTER IX

MULTIPLE SAMPLING AND SEQUENTIAL ANALYSIS

The contents of this proposed chapter are implied by the title. Mixed attributes and variables sampling plans would be included.

Date Due

PRINTED IN U. S. A.

Manufactured in the United States for New York University Press
by the University's Office of Publications and Printing